

CHAPTER II. MANAGEMENT OF WEB SITE AND PORTAL RECORDS

1. OVERVIEW. This chapter establishes policy for the management of both web content and web administrative records. Web content records represent the information presented on a web site and web administrative records provide evidence of the management and operation of web sites. This policy will direct Departmental elements, program staff, webmasters, information technology (IT) staff, and other Department of Energy (DOE) officials who have a role in the management of web sites and portals and the records created in them.
2. BACKGROUND. Managing web site and portal records properly is essential to effective web site operations, especially the mitigation of the risks the Department faces by using the web to carry out its business. Basic assumptions include:
 - a. Reliability, authenticity, integrity and usability of web site and portal information are essential for the Department to meet its legal and business needs and fulfill its missions.
 - i. Reliability means the records can be trusted as accurate.
 - ii. Authenticity means the records were created by the organization with which it is identified.
 - iii. Integrity refers to the fact that the records have not been altered.
 - iv. Usability is based on how easily the web site can be located and retrieved
 - b. For web and portal records to meet these needs, the content, context and structure of the site must be maintained properly.
 - i. Content is the HTML-encoded page or content created by end users interacting with the web site or portal
 - ii. Context refers to the administrative and technical records necessary for or produced during the management of the web sites.
 - iii. Structure is a site map indicating the arrangement of the web site's content pages and its software configuration files.
3. WEB SITE/PORTAL USES. The Department uses the web site/portal in a variety of ways. Records are created when organizations use the web site to conduct program activities. Some of the web site/portal uses include but are not limited to:
 - a. Communications Tool
 - b. Stable Repository for publications (e.g., electronic versions of printed DOE publications)
 - c. Repository for information that changes frequently
 - d. Search or query-based access to information
 - e. Interface to multiple applications and information services

4. WEB SITE/PORTAL MANAGEMENT. A collaborative effort is essential to manage the web site and portal records created. Effective management includes records managers and individuals who have responsibilities relating to web content and web site operations working effectively together. This team effort involves:
 - a. Documentation of web site/portal development
 - b. Procedures for web site/portal operations and records that document those operations
 - c. Coordination of overall presentation of web sites/portals
 - d. Processes and procedures for approval of program-related content materials
 - e. Updates of web/portal sites
 - f. Version controls/tracking of changes
 - g. Reviews for potential legal issues
5. RESPONSIBILITIES. Responsibilities for managing the Department's web site and portal records are coordinated through a team of individuals at different organizational and staff levels. Those individuals are identified throughout this policy. Responsibilities recorded in this policy are at the organizational level.
 - a. OFFICE OF THE CHIEF INFORMATION OFFICER must:
 - i. Establish policy to ensure adequate management and control over Departmental web site and portal records;
 - ii. Ensure that web content and web management and operation records are created, scheduled, and maintained properly;
 - iii. Provide guidance for periodic risk assessments of web sites and portals
 - iv. Provide guidance for the records management aspects of web site/portal operation.
 - b. PROGRAM OFFICES must:
 - i. Approve program-related content material in accordance with Information Quality Guidelines
 - ii. Provide for content creation, graphic design, editorial, technical and other services supporting web site development and operation
 - iii. Coordinate development of web record schedules through their Records Liaison Officers (RLO)
 - iv. Coordinate risk assessments and other IT services supporting web site/portal management and operations with the appropriate webmasters, content managers, and other IT staff
 - c. GENERAL COUNSEL must:

- i. Provide guidance on whether records are trustworthy for legal purposes.
 - ii. Provide legal counsel and advice on or review of potential legal issues such as E-FOIA and incorporation of copyrighted materials in web site and portal design. Additional information on copyright can be found at Copyright.gov.
- 6. REQUIREMENTS. Organizations sponsoring web sites and portals are required to ensure that the information content is an accurate representation of DOE activities and supports the mission of the organization. These records must be protected against unauthorized addition, deletion, and alteration.
 - a. Organizations must manage web site/portal records in accordance with Federal regulations and Departmental policies. It may be necessary for organizations to develop organization-specific requirements to protect the integrity of its records based upon the business, legal, and regulatory requirements of its program mission and organizational DOE-related activities.
 - b. With the support of the IT staff, Program Records Officers (PRO) must conduct risk assessments within their organizations to establish records management controls and documentation requirements for organizational activities. Risk relates to the legal challenge of the trustworthiness of the records and unauthorized loss or destruction of records. Consequences can be measured by the degree of loss that the Department would suffer if the trustworthiness of the web site-related records could not be verified or if there were unauthorized loss or destruction. Risk assessments must be conducted in accordance with this policy. Additional guidance on risk assessment can be obtained from the 'NARA Guidance on Managing Web Records.
 - c. Webmasters and individuals responsible for web content provide support in risk mitigation issues. Webmasters and web content staff shall mitigate risk by:
 - i. producing a web snapshot of all content pages
 - ii. having a site map that shows relationships
 - iii. documenting web site content
 - iv. tracking changes to web sites between snapshots
 - v. ensuring the preservation of hyperlinks when preserving long-term web content pages
 - d. Assisted by the Departmental Records Officer, organizations through their PRO, must develop a plan to address records issues (i.e. types of records needed for documentation, length of retention, security, controls, etc.). The plan should:

- i. Document the systems used to create and maintain web site/portal records
- ii. Ensure web/portal records are created and maintained in secure environments that protect the records from unauthorized alteration or destruction
- iii. Implement standard operating procedures for the creation, use and management of web/portal records and maintain adequate documentation of the procedures
- iv. Document standards, policies and procedures of this Directive for the creation and maintenance of web/portal records.
- v. Develop retention/disposition schedules for web site/portal records and obtain approval of the schedules by the Archivist of the United States (through the OCIO).

7. WEB SITE/PORTAL RECORDS. Web site-related records may exist in electronic or non-electronic format. Each organization may not produce all of these types of records, and there may be other types of records not on the list. This list is not intended to instruct about composition or management of web sites but to provide examples of the type of records that must be managed.

- a. HTML-encoded pages: The content pages composing an agency web site, inclusive of the HTML markup.
- b. Records generated interactively on the web site: Records that are created dynamically in real time when a user interacts with an agency web site (e.g., on-the-fly, text-based page creation, forms filled out online, etc.).
- c. Additional content files referenced within HTML-encoded pages: Files having the ability to "self-execute" (e.g., CGI scripts, Java/ActiveX applets, customized programs that generate online sound or moving images) as well as files that are static (e.g., these include graphic files, multi-national character sets, etc.). Both self-executing and static pages are external to the HTML-encoded content pages but referenced in the HTML syntax.
- d. Comprehensive list of URLs referenced by a web site's hyperlinks: In effect, a bibliography of all uniform resource locators referenced via hyperlinks embedded within a web site's content pages, along with a textual reference describing the destination of the hyperlink.
- e. Web site design records: Records such as graphic design layouts for pages or embedded image maps, and/or minutes of meetings documenting the production of such.
- f. Copyrighted web content and records documenting the use of such content materials: Many Federal web sites incorporate copyrighted content (e.g.,

works for hire such as custom produced graphics files) in individual web pages. Additional records may be deemed necessary for documenting the appropriate use.

- g. Web site program operations records: Program management files that document policies and procedures for agency web site operations, including those that:
 - i. Provide detailed procedures for documenting how records are selected, created and approved for web posting, and how they will be revised or removed
 - ii. Specify what records will be created and how they will be created for interactive sections of web sites
 - iii. Document procedures used in the operation of the site
 - iv. Specify the relationship of the webmaster and other staff involved in preparing and posting web documents to program officials and the agency records officer
 - v. Demonstrate the development of policies and procedures to ensure Section 508 compliance
 - vi. Otherwise explain or illustrate site development and management procedures
 - h. Web site and portal system software-related records: Records related to the application software used in conjunction with operation of the web site. These include:
 - i. Records produced in the analysis and selection of any commercial off-the-shelf (COTS) software
 - ii. Records describing customization of COTS web-related software
 - iii. Documentation relating to COTS web-related software
 - i. Web site logs and statistical compilations: Records including raw data or summary logs of user access (frequency of "hits"), site posting logs (indicating when pages were posted, updated, or removed from the site), system load and search result statistics, and ad hoc reports containing such.
 - j. Web site map: A linked, graphic or text-based display of a web site's hierarchy, similar to an organization chart.
 - k. COTS software configuration files: Files used to operate the web site and establish its look and feel including server environment configuration specifications.
8. RISK AND RISK ASSESSMENTS. Organizations must periodically perform risk assessments of its web site and portal records management control and operational needs. The assessment will establish the appropriate levels of internal management control required to determine which records are necessary to ensure

reliability, authenticity, integrity; how the records should be maintained; and the length of time the records should be retained. The assessment should include the nature and sensitivity of the information on the site/portal, who uses the information, and what problems might arise if the information is incorrect, out of date, or lost. It should also provide documentation on the frequency of update and the security required to maintain the trustworthiness of the web site/portal records.

- a. Risks. Many records management-related risks associated with web sites are technical risks which could result in negative programmatic consequences:
 - i. Inability to document or validate transactions
 - ii. Inability to reconstruct views of web/portal content
 - iii. Compromise of information
 - iv. Inability to track web/portal-assisted policy development or document agency decisions relating to web operations
 - v. Inability to verify site information at any given time; especially important for litigation
 - vi. Impair program operations
 - vii. Inability to detect or punish fraud, false statements, or other illegal behavior because of lack of valid records
 - viii. Inability to produce records for accountability and stewardship of information
 - ix. Dissemination of misinformation; non-compliance with Information Quality Guidelines
 - x. Compromise of citizens/government rights or agency missions
 - xi. Unfavorable media attention; negative reaction of stakeholders.
- b. Assessments. A records management risk assessment shall address the possible consequences of untrustworthy, lost, or unrecoverable records, including the legal risk and financial costs of losses, the likelihood that a damaging event will occur, and the costs of taking corrective actions. The assessment factors should include records management threats, visibility, consequences, and sensitivity.
 - i. Records management threats include risks of unauthorized destruction of web site/portal records, the inability to reconstruct views of the web sites for litigation purposes, and the inability to document web site policy decisions
 - ii. Visibility is the level of public awareness of the web site/portal
 - iii. Consequences describes the level of negative organizational, economic, or programmatic impact if web site/portal records are untrustworthy, lost, or unrecoverable
 - iv. Sensitivity relates to the Department's assessment of the importance of the web site/portal records and its operations.
- c. Conducting an Assessment. In conducting an assessment, consider whether the web site has a single level of risk or varying levels of risk

through the use of risk assessment factors. If the level chosen for analysis has more than one answer to any of the factors, consider breaking out those portions. Changes in any four factors could affect the risk level.

- i. Evaluate the web site in toto. Note that this option is not advisable if the web site has multiple types of content (e.g., e-commerce transactions and static publications) or functions served. Records management risk and required management controls vary for those different portions of the web site
- ii. Evaluate groupings of web sites referenced by the main portal entry page
- iii. Evaluate the web site basically as a whole, minus one or two portions that exhibit substantially different characteristics
- iv. Substantially break out clusters or groups of web site pages based on function or other characteristics. Note that this option does not anticipate a page-by-page risk analysis of your web site.

If for operational reasons, the web site is evaluated as a single unit, all components must be treated the same in terms of risk. Manage all parts of the site in accordance with the highest level of risk determined for any portion of the site. When applying this policy to portals that are primarily federations of the Department's web sites, manage the web sites at the highest level of risk encountered in the aggregation.

- d. Mitigating Risk. Organizations must protect the records appropriately after determining the level of records management risks for their web sites/portals or portions thereof. The results of the risk assessment will indicate the level of effort necessary to mitigate risk.
 - i. Review any web management policies and procedures already in place to determine whether additional steps are needed
 - ii. Develop a plan to address records issues (e.g., types of records needed to document the web-based activity, length of time they are needed to support the business purposes), as well as IT issues (e.g., security of the site and information exchanged over the site) and management/internal controls on the processes
 - iii. Document the systems used to create and maintain your web records
 - iv. Ensure that your web records are created and maintained in a secure environment that protects the records from unauthorized alteration or destruction
 - v. Implement standard operating procedures for the creation, use, and management of your web records and maintain adequate written documentation of those procedures
 - vi. Create and maintain web records according to documented standard operating procedures
 - vii. Train staff in the standard operating procedures

- viii. Develop a retention schedule for the web records and obtain official NARA approval of that retention schedule. Cite the official disposition authorities found in the DOE Administrative Schedules if there are legal challenges to produce records that have been destroyed
- e. Management of Web Site Content. Web content pages may be frequently changed or updated resulting in a change in relation/organization of web content. If a change or redesign occurs, it will be necessary to preserve the change/redesign as a content web record. Organizations have the option to create the new web site record through the use of a snapshot or a content management system (CMS). A snapshot captures a web site as it existed at a particular point in time. If the content records are permanent, a method that retains hypertext functionality is to be used. A CMS keeps the website updated by managing the creation, modification, and removal of content from the website.
 - i. If the snapshot function is employed, take a snapshot of all content pages on the site at a particular time and include a site map that shows the relationship of those pages to each other. Decide how frequently a new snapshot should be captured, if it is necessary to track changes in both the content pages and the site map that occur between snapshots, and how to track those changes if they occur.
 - ii. If a CMS is used, it should support the creation of audit trails. Web server backup capability or Internet-based service (i.e., a “mirror” of the original site) must be available in case of equipment failure or other catastrophes.
 - iii. An option available for low risk web sites only is to manage the live versions of the web site content pages while the pages are up on the web site. The current posted version of the web site and the standard operating procedures (SOP) and a log of changes may be sufficient. This option is not recommended for medium- and high-risk sites.
- f. Tracking Changes to Mitigate Risks. Organizations must decide how to best keep track of the versions of their content pages. The initial risk assessment will be instrumental in determining what approach is acceptable for version control issues.
 - i. CMS software can be used to track versions of web content.
 - ii. Major changes to the site’s directory structure can be handled by producing a new site map at the time of major revision.
 - iii. A DoD 5015.2-certified records management application (RMA) used to manage web content records can be utilized to track versions of web content pages. A DoD certification means that the RMA complies with the NARA records management regulations.
- g. Hyperlink Management for Long Term Preservation. It is possible that hyperlinks will not continue to function in the preservation copy of the

web content records. If the site does not follow external-link-liability transference policies (for example, pop-up window notifications) organizations must consider the following:

- i. For hyperlinks that simply send the user to a different location within the same page (aka internal target), no additional work is required, as the link will continue to function when the content page is interpreted by a browser application.
- ii. For hyperlinks that send the user to either a different page or another web site that is not under the agency's records management control, web site content developers are required to modify the HTML syntax of web content pages containing such hyperlinks on a day-forward basis. Modification would include the insertion of an HTML comment after the hyperlink that describes the name of the site and perhaps a content summary or page to which the hyperlink transfers.
- iii. When a page includes a hyperlink that sends the user to another page in the same web site, insert a comment describing the hyperlink is not being scheduled in toto for the same retention. The comment should reference the series containing the destination of the hyperlink.
- iv. Another alternative would be to produce what is in effect a bibliography for all of the hyperlinks referenced within the content pages composing a site. List all of the URLs referenced by hyperlinks, along with a description of the hyperlinked page (much as in the comment used in the previously suggested method).

9. SCHEDULING WEB SITE /PORTAL RECORDS. A web schedule covers the web content records that document the information on the site and the web site management and operations records that provide the site's context and structure. A web schedule is important because it fulfills statutory records management responsibilities, mitigates risks associated with websites/portals, and provides legal authority to destroy web records based on NARA-approved retention periods. There are currently no schedules in NARA's General Records Schedules (GRS) or the Department's Administrative Schedules that were developed to specifically cover web records. However, there may be some items in DOE Administrative Schedules 20 and 24 that may be applied to web site management and operations records. The organization's PRO must coordinate the development of unique web schedules through the Departmental Records Officer. Organizations can exercise several options in scheduling web site/portal records. In addition to this policy, see guidance for scheduling web records in "NARA Guidance on Managing Web Records."

- a. Single schedule for all web/portal records. Use a single schedule for content and management/operating records if all the records related to the site warrant the same retention period in order to meet business needs and mitigate risks.

- b. Single schedule item for web content and single item for web management and operating records. A separate schedule for content and management/operating records allows for development of different levels of retention to mitigate risks.
- c. Multiple schedules for all web/portal records. If it is warranted, schedule web content records separately from web management/operating records.
 - i. A single schedule item is appropriate for management/operating records if the records are needed for the same period of time in order to mitigate risk. Web content records can be covered by one or more separate schedules.
 - ii. If records are needed for different periods of time in order to mitigate risk, then group the records together according to their risk level and assign the appropriate retention schedule. Multiple schedules should then be developed for the management/operations records, regardless of whether web content records are included in a single schedule or multiple schedules.
- c. Web Snapshots. The schedule for a snapshot shall include snapshots that capture the content pages and related site map as they existed at particular points in time. Business requirements and risk mitigation determine if snapshots and the frequency of them are warranted.

REFERENCES:

44 U.S.C., Chapter 33 Disposal of Records

http://www.law.cornell.edu/uscode/html/uscode44/usc_sup_01_44_10_31.html

44 U.S.C., Chapter 35 Coordination of Federal Information Policy

http://www.law.cornell.edu/uscode/html/uscode44/usc_sup_01_44_10_35.html

44 U.S.C., 3101 Records Management By Federal Agencies

http://www.law.cornell.edu/uscode/html/uscode44/usc_sec_44_00003101----000-.html

E-Government Act of 2002 (Public Law 107-347) – Enhances the management and promotion of electronic Government services and processes by establishing a broad framework of measures that require using Internet-based information technology to enhance citizen access to Government information and services, and for other purposes.

<http://www.regulations.gov/images/e-Gov%20Law.pdf>

DoD 5015.2-STD, Design Criteria Standard for Electronic Records Management Software Applications – A standard developed by the Defense Department and endorsed by NARA that sets forth mandatory baseline functional requirements for Records Management Application software, defines required system interfaces and search criteria to be supported, and describes the minimum records management requirements that must be met based upon current NARA regulations.

<http://www.dtic.mil/whs/directives/corres/html/50152std.htm>

DOE-STD-4001-2000, Design Criteria Standard for Electronic Records Management Software Applications – DOE 'S adaptation of DoD 5015.2

<http://cio.doe.gov/RBManagement/Records/poliguide.html>

NARA Guidance on Managing Web Records

http://www.archives.gov/records_management/index.html

NARA Guidance on Transferring Permanent Web Records

http://www.archives.gov/records_management/initiatives/web_content_records.html

DEFINITIONS:

Content – Any material that is available on a Federal public web site or portal.

Content Management System – Software that facilitates web site administration by providing suites of web-related functionality that may include template or web content construction, on the fly page creation from databases, versioning control, workflow and import/export features.

Context – The organizational, functional, and operational circumstances in which documents are created and/or received and used.

HTML – Hypertext Markup Language is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page.

Hyperlink – The HTML syntax for expressing the ability for one internet document to reference another document located either in another portion of the web site or at another web site on the WWW.

Portal – A term generally synonymous with gateway for a web site that is or proposes to be a major starting site for users when they get connected to the web or that users tend to visit as an anchor site.

Records Schedule – A document describing records of an agency, organization, or administrative unit, establishing a timetable for their life cycle, and providing authorization for their disposition, i.e., destruction or transfer to the National Archives.

Records Management Application – Software used by an organization to manage its records. Primary management functions are categorizing and locating records and identifying records that are due for disposition. Also stores, retrieves and disposes of the electronic records that are stored in its repository.

Structure – The physical and logical format of a record and the relationships between the data elements.

Site Map – A linked, graphic or text-based display of a web site's hierarchy, similar to an organization chart. Typically, site maps break down a web site's content into increasingly specific subject areas to help the visitor understand its structure, from the main navigation pages to their subordinate pages.

Web Site Administrative Records – Federal records generated by the web site program office in the course of its operations.

Web Site Content Records – Web site content determined by the agency to meet the statutory criteria in 44 USC 3301 for being deemed a Federal record.